RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 10/564,560
Source: IFW
Date Processed by STIC: 1-30-06

ENTERED

IFWP

RAW SEQUENCE LISTING DATE: 01/30/2006
PATENT APPLICATION: **US/10/564,560** TIME: 15:01:07

Input Set : **A:\seq listing.app**
Output Set: **N:\CRF4\01302006\J564560.raw**

```
      3 <110> APPLICANT: Graham and Tonon
      5 <120> TITLE OF INVENTION: Transgenic Cells
      7 <130> FILE REFERENCE: 72576-01
C-->  9 <140> CURRENT APPLICATION NUMBER: US/10/564,560
C-->  9 <141> CURRENT FILING DATE: 2006-01-12
      9 <150> PRIOR APPLICATION NUMBER: PCT/GB04/003057
     10 <151> PRIOR FILING DATE: 2004-07-13
     12 <150> PRIOR APPLICATION NUMBER: 0316629.5
     13 <151> PRIOR FILING DATE: 2003-07-16
     15 <160> NUMBER OF SEQ ID NOS: 19
     17 <170> SOFTWARE: PatentIn version 3.1
     19 <210> SEQ ID NO: 1
     20 <211> LENGTH: 1702
     21 <212> TYPE: DNA
     22 <213> ORGANISM: Thalassiosira
     24 <220> FEATURE:
     25 <221> NAME/KEY: misc_feature
     26 <222> LOCATION: (1)..(1)
     27 <223> OTHER INFORMATION: undefined nucleotide base
     30 <220> FEATURE:
     31 <221> NAME/KEY: misc_feature
     32 <222> LOCATION: (1700)..(1700)
     33 <223> OTHER INFORMATION: undefined nucleotide base
     36 <220> FEATURE:
     37 <221> NAME/KEY: misc_feature
     38 <222> LOCATION: (3)..(3)
     39 <223> OTHER INFORMATION: undefined nucleotide base
     42 <220> FEATURE:
     43 <221> NAME/KEY: misc_feature
     44 <222> LOCATION: (1701)..(1701)
     45 <223> OTHER INFORMATION: undefined nucleotide base
     48 <400> SEQUENCE: 1
W--> 49 ncncgcagca gggtgcgcgc aaattttgaa ggaattttcg agggtgatta gtgcggcgac   60
     51 taagtgcgga tgacaataaa gattaacggg atttatcaac ggggaaaaca cggaaatatg  120
     53 ttcccgtaga gaaataatgg tctcgttaaa ttagttataa tcttaaacaa tttagtggtc  180
     55 aatataatag acaaaaatga caaattagtg gactctttcg ccaccaactc ttcacaagac  240
     57 caactgtttt tgtgccctcc tccctctcag ttgctacgat tcgctgacct cctttctcta  300
     59 ctaccgtcgg ctccaacacc atcatcatgt cgcaattcct aaccagcatc cccaaggaat  360
     61 gcgtaggcac caacggcctc ggagtccact acgccgaatt ctcctgcctc caccctctcc  420
     63 tcggcgccac ctacctcccc ttcgaacgct tctacgatcc cgtcgccacc ctcacctgga  480
     65 tgcaagatcg tcccatgatc cccatcatcg cctgcgtcgc ctacgtcgtg ctcatcgtcc  540
     67 tgggacgcgc ctacatgaag gaccggccgg cgtggagctg gaggaggatt ttggccgttt  600
     69 ggaatttgag cctgtcgctc ttctcgtgga ttggcgcgat caggacggct cctcagttgt  660
```

RAW SEQUENCE LISTING DATE: 01/30/2006
PATENT APPLICATION: US/10/564,560 TIME: 15:01:07

Input Set : A:\seq listing.app
Output Set: N:\CRF4\01302006\J564560.raw

71 attacaactt gacgacgtat tcgttgaggg ataatttgtg cgatgatccg gcggcgttgt 720
73 atgggagtgg atcgacggga ctttgggtgc agttgttcat tttgagcaag tttcccgagt 780
75 tgcttgacac tttcttcatt gtcattcaca agaagccgct catcttcctc cattggtatc 840
77 atcatatcac cgtccttctt tactgctggc attcctatgt gaccacttct cccagtggtc 900
79 tcttcttcgt cgtcatgaac tacagtgtcc acgcggtcat gtatgggtac tacttcctca 960
81 tggcggtcaa attccgtccc aaatggttca accccatgtt cgtgacgttc atgcaacttt 1020
83 ctcaaatgtt tattggggtg ggagttacca ttgtggcatt ttattattac agtaatccga 1080
85 ttttgggaaa gacatgtcat atcaggaagg agaacaatgt tgcggccttt gtcatgtacg 1140
87 ggagctactt ttacttgttt gcacaattct ttgtggcgag gtattataag gttaaggtca 1200
89 agggggatgc gaagaagaag aaggttgtgt aaagtgagag atggaatgaa acaaccatct 1260
91 tgtttgggga agggggtatt ggatagcggg taccattcag tatcgttgag gtgcatttaa 1320
93 tgttgaatga acaaacttga cgagacgagg gattttgatc ttcatgaacg agtgggagca 1380
95 tctttcaatc cattggggag agaggagaag tgagagaagt gctactttgg gagtttgaga 1440
97 gagtaaatta atgtcttttg ctatgaattg ctgcctcaaa aacgcaacgt gctagcaaac 1500
99 ctcgttaaca atgacaaagt tatttcttgt tgtatgggac ataccacgat tgtatcataa 1560
101 aagaaaacca attctattga gttgtaaaca tctagagtgc agtatcgagc aacagcccac 1620
103 gccatcacga tacactaaac acacattcgt cttcatcttt acattctaac cacagcatgc 1680
105 tggctctctt acctcttcan nc 1702
108 <210> SEQ ID NO: 2
109 <211> LENGTH: 1224
110 <212> TYPE: DNA
111 <213> ORGANISM: Thalassiosira
113 <400> SEQUENCE: 2
114 acgcggggtt gctacgattc gctgacctcc tttctctact accgtcggct ccaacaccat 60
116 catcatgtcg caattcctaa ccagcatccc caaggaatgc gtaggcacca acggcctcgg 120
118 agtccactac gccgaattct cctgcctcca ccctctcctc ggcgccacct acctcccctt 180
120 cgaacgcttc tacgatcccg tcgccaccct cacctggatg caagatcgtc ccatgatccc 240
122 catcatcgcc tgcgtcgcct acgtcgtgct catcgtcctg ggacgcgcct acatgaagga 300
124 ccggccggcg tggagctgga ggaggatttt ggccgtttgg aatttgagcc tgtcgctctt 360
126 ctcgtggatt ggcgcgatca ggacggctcc tcagttgtat tacaacttga cgacgtattc 420
128 gttgagggat aatttgtgcg atgatccggc ggcgttgtat gggagtggat cgacgggact 480
130 ttgggtgcag ttgttcattt tgagcaagtt tcccgagttg cttgacactt tcttcattgt 540
132 cattcacaag aagccgctca tcttcctcca ttggtatcat catatcaccg tccttcttta 600
134 ctgctggcat tcctatgtga ccacttctcc cagtggtctc ttcttcgtcg tcatgaacta 660
136 cagtgtccac gcggtcatgt atgggtacta cttcctcatg gcggtcaaat tccgtcccaa 720
138 atggttcaac cccatgttcg tgacgttcat gcaactttct caaatgttta ttggggtggg 780
140 agttaccatt gtggcatttt attattacag taatccgatt ttgggaaaga catgtcatat 840
142 caggaaggag aacaatgttg cggcctttgt catgtacggg agctactttt acttgttcgc 900
144 acaattcttt gtggcgaggt attataaggt taaggtcaag ggggatgcga agaagaagaa 960
146 ggttgtgtaa agtgagagat ggaatgaaac aaccatcttg tttggggaag ggggtattgg 1020
148 atagcgggta ccattcagta tcgttgaggt gcatttaatg ttgaatgaac aaacttgacg 1080
150 agacgaggga ttttgatctt catgaacgag tgggagcatc tttcaatcca ttggggagag 1140
152 aggagaagtg agagaagtgc tactttggga gtttgagaga gtaaattaac gtcttttgca 1200
154 aaaaaaaaaa aaaaaaaaaa aaaa 1224
157 <210> SEQ ID NO: 3
158 <211> LENGTH: 1828
159 <212> TYPE: DNA
160 <213> ORGANISM: Thalassiosira
162 <220> FEATURE:

RAW SEQUENCE LISTING DATE: 01/30/2006
PATENT APPLICATION: **US/10/564,560** TIME: 15:01:07

Input Set : **A:\seq listing.app**
Output Set: **N:\CRF4\01302006\J564560.raw**

```
     163 <221> NAME/KEY: misc_feature
     164 <222> LOCATION: (1824)..(1824)
     165 <223> OTHER INFORMATION: unknown nucleotide
     168 <400> SEQUENCE: 3
     169 gtggtctcat ggcgtgggtc gcttggttct ccttcgctgt gctcccctct ctccttctcg     60
     171 gcggtgtgtg cggtcttcgt tttcatttgc cttcttttcc catcaggttt cctagacgtg    120
     173 cggggccgcg tccttctctt gggttgggct tgcccgcttt ggtttgatat cacaacagtt    180
     175 acctggcaac catggacgct tacaacgctg caatggataa gatcggtgcc gccatcatcg    240
     177 attggtctga tcccgatggc aagttccgtg ccgatagaga ggtgagcatg aatgtacaca    300
     179 ccatggttgt ctcggcatga cggtgtcatt ggatggtgtc agtgcatctc tctgtttgca    360
     181 tctattctaa acaacacatc tcttcacctc gttaccttac tcaacaacta ccacacaacc    420
     183 atcatcatcg taggactggt ggctctgcga cttccgtagc gccatcacca tcgctctcat    480
     185 ctacatcgcc ttcgtcatcc tcggttccgc cgtcatgcaa tccctccccg caatggatcc    540
     187 ctaccccatc aaattcctct acaacgtctc ccaaatcttc ctttgtgcct acatgactgt    600
     189 cgaggcggga tttttggcct accgcaatgg atataccgtc atgccttgca atcatttcaa    660
     191 tgtgaatgat cctcccgtgg cgaatcttct ttggttgttt tatatttcca aggtgtggga    720
     193 cttttgggat accattttca ttgtgttggg gaagaagtgg cgtcaattat ctttcttgca    780
     195 tgtataccat cacaccacca tctttctatt ctattggctg aatgccaatg tcttgtacga    840
     197 tggtgacatc ttccttacca tcttgctcaa tggattcatc cacacggtga tgtacacgta    900
     199 ttacttcatc tgtatgcata ccaaagatcc caagacgggc aagagtcttc ctatatggtg    960
     201 gaagtcgagt ttgacggcgt ttcagttgtt gcaattcact atcatgatga gtcaggctac   1020
     203 ctaccttgtc ttccacgggt gtgataaggt gtcgcttcgt atcacgattg tgtactttgt   1080
     205 gtacattttg agtttgttct tccttttgc tcagttcttt gtgcaatcat acatggcacc   1140
     207 caaaaagaag aagagtgctt agattggaaa ggggtgtggg cgacgagctt tcctgttgag   1200
     209 ggtgggtggt ggaacggagt tggtttttg aagcatctgc aatattcgca ggactgttgc   1260
     211 tgtgagaata gctatggagt aaaggtgggg gggggtggat tcatggcgga caggcatgcc   1320
     213 taagatacta aggaatgttc atgaacatga tgttgatact ttattgtaag gtactgttgg   1380
     215 gaattaatga gagggtactg aaaggagaga tgagtgtctg tcaaaacgct tgggttagtt   1440
     217 gttactttcc cttcgttctt tcagctataa gtctttgctg aggagttaat cctaagctga   1500
     219 caccattacg ttgaacaacg caacaattag cgttgagccc gacaactctc gacaaagagg   1560
     221 ttttgtagat tgtatcccct ggcgcaagtt aacgtacagg tccttcattc acggaaccat   1620
     223 aatcccatgg atgcatcctg tgccaataac cttcaaaaac gtgcgtccca cttgagaaaa   1680
     225 ccactattac gagtttcacc tcaggtcctg accggcaaaa acaattgaat cagcagcaaa   1740
     227 gcccacaagc aagcacttcg cgatgaggac cacaggaaga gacgctcaca cctccccgcc   1800
W--> 229 ttcggacgag ccccacgagc gcgngtgg                                    1828
     232 <210> SEQ ID NO: 4
     233 <211> LENGTH: 1566
     234 <212> TYPE: DNA
     235 <213> ORGANISM: Thalassiosira
     237 <220> FEATURE:
     238 <221> NAME/KEY: misc_feature
     239 <222> LOCATION: (25)..(25)
     240 <223> OTHER INFORMATION: unknown nucleotide
     243 <220> FEATURE:
     244 <221> NAME/KEY: misc_feature
     245 <222> LOCATION: (1564)..(1564)
     246 <223> OTHER INFORMATION: unknown nucleotide
     249 <400> SEQUENCE: 4
W--> 250 aaaaaaaaaa aaaaaaaaaa aaganaggaa atgtcgcaca acggcagctg cagcttcatg     60
```

RAW SEQUENCE LISTING DATE: 01/30/2006
PATENT APPLICATION: **US/10/564,560** TIME: 15:01:07

Input Set : **A:\seq listing.app**
Output Set: **N:\CRF4\01302006\J564560.raw**

```
252 cctgcaggtc gactctagag gatccccgtg ttgtcaatgt ggcgcaagtg ctgctcaaat    120
254 gggtggacgg tgtatgcgat tgtggatgcg gtgatgaata gagaccatcc atttattgga    180
256 agtagaagtt tggttggggc ggcgttgcat agtgggagct cgtatgcggt gtgggttcat    240
258 tattgtgata agtatttgga gttctttgat acgtatttta tggtgttgag ggggaaaatg    300
260 gaccaggtga gttgacgagt tgctgtttag tgttggttag atggtacttg gtgaagttgg    360
262 tgacagtgtg tggtgtggcg ttggatatat ggatatggag aaggtaccaa ttggttggaa    420
264 ggaacaatga gacacatcct gcgcacagtg tccagagaga cgaatctgca acgattcaaa    480
266 gatcatttaa gagttcatca gctacgcaga agaatggtgt aaatgatact gttcagtttc    540
268 aaaagttggc atgatactac tcagctttga agtgcatcgg tctgctcagg gaacggggaa    600
270 ggctttcacc aacaacgtta ccaatccaca tctcacgctt ccacctcatt ctacaaaaac    660
272 aaaaaaacag gtctccttcc tccacatcta ccaccacacg accatagcgt gggcatggtg    720
274 gatcgccctc cgcttctccc ccggcggaga catttacttc ggggcactcc tcaactccat    780
276 catccacgtc ctcatgtatt cctactacgc ccttgcccta ctcaaggtca gttgtccatg    840
278 gaaacgatac ttgactcaag ctcaattatt gcaattcaca agtgtggtgg tttatacggg    900
280 gtgtacgggt tatactcatt actatcatac gaagcatgga gcggatgaga cacagcctag    960
282 tttaggaacg tattatttct gttgtggagt gcaggtgttt gagatggtta gtttgtttgt   1020
284 actcttttcc atcttttata aacgatccta ttcgaagaag aacaagtcag gaggaaagga   1080
286 tagcaagaag aatgatgatg ggaataatga ggatcaatgt cacaaggcta tgaaggatat   1140
288 atcggagggt gcgaaggagg ttgtggggca tgcagcgaag gatgctggaa agttggtggc   1200
290 tacggcgagt aaggctgtaa agaggaaggg aactcgtgtt actggtgcca tgtagataaa   1260
292 gaggttgaag agagatgaag gcaactcttc atgatggtgg tcgaagtttc atcaacatta   1320
294 actgtatgaa tcaagataaa ggtggttgga caaggatgtc ttggaatacg gcatgaatag   1380
296 gagaacaagt tgctaatgat tctagagaat gtacattcag acttcgtgta taaagacgat   1440
298 actccgggat cgtcacgtac cgtttcgaag taggccatgc tcaagaccgt gatatactga   1500
300 gtgcgccgat ctatctactt gaagccatcc ttactgtgcg gcgatcgaac aagaattccc   1560
302 gacngg                                                             1566
305 <210> SEQ ID NO: 5
306 <211> LENGTH: 301
307 <212> TYPE: PRT
308 <213> ORGANISM: Thalassiorsira
310 <400> SEQUENCE: 5
312 Met Ser Gln Phe Leu Thr Ser Ile Pro Lys Glu Cys Val Gly Thr Asn
313 1               5                  10                  15
316 Gly Leu Gly Val His Tyr Ala Glu Phe Ser Cys Leu His Pro Leu Leu
317             20                  25                  30
320 Gly Ala Thr Tyr Leu Pro Phe Glu Arg Phe Tyr Asp Pro Val Ala Thr
321         35                  40                  45
324 Leu Thr Trp Met Gln Asp Arg Pro Met Ile Pro Ile Ile Ala Cys Val
325     50                  55                  60
328 Ala Tyr Val Val Leu Ile Val Leu Gly Arg Ala Tyr Met Lys Asp Arg
329 65                  70                  75                  80
332 Pro Ala Trp Ser Trp Arg Arg Ile Leu Ala Val Trp Asn Leu Ser Leu
333                 85                  90                  95
336 Ser Leu Phe Ser Trp Ile Gly Ala Ile Arg Thr Ala Pro Gln Leu Tyr
337             100                 105                 110
340 Tyr Asn Leu Thr Thr Tyr Ser Leu Arg Asp Asn Leu Cys Asp Asp Pro
341         115                 120                 125
344 Ala Ala Leu Tyr Gly Ser Gly Ser Thr Gly Leu Trp Val Gln Leu Phe
345     130                 135                 140
```

RAW SEQUENCE LISTING DATE: 01/30/2006
PATENT APPLICATION: US/10/564,560 TIME: 15:01:07

Input Set : A:\seq listing.app
Output Set: N:\CRF4\01302006\J564560.raw

```
348 Ile Leu Ser Lys Phe Pro Glu Leu Leu Asp Thr Phe Phe Ile Val Ile
349 145                 150                 155                 160
352 His Lys Lys Pro Leu Ile Phe Leu His Trp Tyr His His Ile Thr Val
353                 165                 170                 175
356 Leu Leu Tyr Cys Trp His Ser Tyr Val Thr Thr Ser Pro Ser Gly Leu
357             180                 185                 190
360 Phe Phe Val Val Met Asn Tyr Ser Val His Ala Val Met Tyr Gly Tyr
361         195                 200                 205
364 Tyr Phe Leu Met Ala Val Lys Phe Arg Pro Lys Trp Phe Asn Pro Met
365     210                 215                 220
368 Phe Val Thr Phe Met Gln Leu Ser Gln Met Phe Ile Gly Val Gly Val
369 225                 230                 235                 240
372 Thr Ile Val Ala Phe Tyr Tyr Tyr Ser Asn Pro Ile Leu Gly Lys Thr
373                 245                 250                 255
376 Cys His Ile Arg Lys Glu Asn Asn Val Ala Ala Phe Val Met Tyr Gly
377             260                 265                 270
380 Ser Tyr Phe Tyr Leu Phe Ala Gln Phe Phe Val Ala Arg Tyr Tyr Lys
381         275                 280                 285
384 Val Lys Val Lys Gly Asp Ala Lys Lys Lys Lys Val Val
385     290                 295                 300
388 <210> SEQ ID NO: 6
389 <211> LENGTH: 242
390 <212> TYPE: PRT
391 <213> ORGANISM: Thalassiorsira
393 <400> SEQUENCE: 6
395 Asp Trp Trp Leu Cys Asp Phe Arg Ser Ala Ile Thr Ile Ala Leu Ile
396 1               5                   10                  15
399 Tyr Ile Ala Phe Val Ile Leu Gly Ser Ala Val Met Gln Ser Leu Pro
400             20                  25                  30
403 Ala Met Asp Pro Tyr Pro Ile Lys Phe Leu Tyr Asn Val Ser Gln Ile
404         35                  40                  45
407 Phe Leu Cys Ala Tyr Met Thr Val Glu Ala Gly Phe Leu Ala Tyr Arg
408     50                  55                  60
411 Asn Gly Tyr Thr Val Met Pro Cys Asn His Phe Asn Val Asn Asp Pro
412 65                  70                  75                  80
415 Pro Val Ala Asn Leu Leu Trp Leu Phe Tyr Ile Ser Lys Val Trp Asp
416                 85                  90                  95
419 Phe Trp Asp Thr Ile Phe Ile Val Leu Gly Lys Lys Trp Arg Gln Leu
420             100                 105                 110
423 Ser Phe Leu His Val Tyr His His Thr Thr Ile Phe Leu Phe Tyr Trp
424         115                 120                 125
427 Leu Asn Ala Asn Val Leu Tyr Asp Gly Asp Ile Phe Leu Thr Ile Leu
428     130                 135                 140
431 Leu Asn Gly Phe Ile His Thr Val Met Tyr Thr Tyr Tyr Phe Ile Cys
432 145                 150                 155                 160
435 Met His Thr Lys Asp Pro Lys Thr Gly Lys Ser Leu Pro Ile Trp Trp
436                 165                 170                 175
439 Lys Ser Ser Leu Thr Ala Phe Gln Leu Leu Gln Phe Thr Ile Met Met
440             180                 185                 190
```

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 01/30/2006
PATENT APPLICATION: **US/10/564,560** TIME: 15:01:08

Input Set : **A:\seq listing.app**
Output Set: **N:\CRF4\01302006\J564560.raw**

Please Note:
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:1; N Pos. 1,3,1700,1701
Seq#:3; N Pos. 1824
Seq#:4; N Pos. 25,1564

VERIFICATION SUMMARY DATE: 01/30/2006
PATENT APPLICATION: **US/10/564,560** TIME: 15:01:08

Input Set : **A:\seq listing.app**
Output Set: **N:\CRF4\01302006\J564560.raw**

L:9 M:270 C: Current Application Number differs, Replaced Current Application No
L:9 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:49 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1 after pos.:0
M:341 Repeated in SeqNo=1
L:229 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3 after pos.:1800
L:250 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4 after pos.:0
M:341 Repeated in SeqNo=4